Project report final

Applied data science

Yuxuan Wang

Date May.9th

Introduction

Resource of Data

The dataset I want to use is 2017-2018 NBA player statistic per game.
The link of data is https://www.basketball-reference.com/leagues/NBA_2018_per_game.html
I have downloaded the csv file.
The other dataset is NBA player salary dataset in season 2017-2018 downloaded from Kaggle, the link is https://www.kaggle.com/koki25ando/salary

Summary for NBA: ([1]NBA.com)

The National Basketball Association (NBA) is a men's professional basketball league in North America; composed of 30 teams (29 in the United States and 1 in Canada). It is widely considered to be the premier men's professional basketball league in the world. (WIKI) There are many detailed statistic and various kinds of records in the league. To estimate the player performance per game, we have points in the game, field goal rate, rebound, steals and so on, which is listed in the dataset.

Current method of calculate players' performance per game: (efficiency)

NBA.com evaluates all players based on the efficiency formula: ((Points + Rebounds + Assists + Steals + Blocks) - ((Field Goals Att. - Field Goals Made) + (Free Throws Att. - Free Throws Made) + Turnovers)). ([2] NBA.com)
The coaches in each team often use this formula to quickly evaluate a player's performance in the game.

Data analysis discussion:

Through the data, what I want to achieve is that :

1. Help coach find promising player easily as Money ball. Especially when team manager does not have enough budgets. For today's league, All star player tends have a really high salary. Due to the salary cap, most of the teams need to find the best role players within limited budgets. An accurate estimation formula should be attractive to team managers to use the team budget more reasonably.
2. I will add players' salary as one of the factor to estimate the performance of a player. Also, I would like to watch more game videos to see there would be more parameters to evaluate a role

player

3. Most important! I will test if there is a linear regression relationship between role players' salary and efficiency. If there is, then how to increase their efficiency and which stats factors have influence on the efficiency like minutes per game or how many field goal attempts per game. And finally, for these role players, how to change their play styles.

Data Issue

1. I have not found the salary for each player in the league. The salary might need to stick with this data to analyze #4. Or I will only focus on middle class players.
2. Due to the complexity of player's contract and salary cap. Some role players with great ability might not own high salary, vice versa.
3. Some player with very few playing chance. How to adjust Their performance data.
4. When there is a huge score gap between 2 teams before the end of game. Coach tends to give role players chances to play. During this period, players usually do not put so much attention and they may easily get many points. How to reduce this kind of error. Also, there will be more trash time in a strong team's game.

Reference

[1] http://www.nba.com/news/faq
[2] http://www.nba.com/statistics/efficiency.html